

What is claimed is:

1. A singing voice synthesizing apparatus comprising:

- 2 a phoneme database that stores a plurality of voice
5 fragment data formed of voice fragments each being a
single phoneme or a phoneme chain of at least two
concatenated phonemes, each of the plurality of voice
fragment data comprising data of a deterministic
component and data of a stochastic component;
10 an input device that inputs lyrics;
a readout device that reads out from said phoneme
database the voice fragment data corresponding to the
inputted lyrics;
a duration time adjusting device that adjusts time
15 duration of the read-out voice fragment data so as to
match a desired tempo and manner of singing;
an adjusting device that adjusts the deterministic
component and the stochastic component of the read-out
voice fragment so as to match a desired pitch; and
20 a synthesizing device that synthesizes a singing
sound by sequentially concatenating the voice fragment
data that have been adjusted by said duration time
adjusting device and said adjusting device.
2. A singing voice synthesizing apparatus according
25 to claim 1, wherein said phoneme database stores a
plurality of voice fragment data having different musical
expressions for a single phoneme or phoneme chain.
3. A singing voice synthesizing apparatus according
to claim 2, wherein said musical expressions include at
30 least one parameter selected from the group consisting of
pitch, dynamics and tempo.
4. A singing voice synthesizing apparatus according
to claim 1, wherein said phoneme database stores voice
fragment data comprising elongated sounds that are each
35 enunciated by elongating a single phoneme, voice fragment

2025-12-27 10:34:59

data comprising consonant-to-vowel phoneme chains and vowel-to-consonant phoneme chains, voice fragment data comprising consonant-to-consonant phoneme chains, and voice fragment data comprising vowel-to-vowel phoneme chains.

5 5. A singing voice synthesizing apparatus according to claim 1, wherein each of said voice fragment data comprises a plurality of data corresponding respectively to a plurality of frames of a frame string formed by
10 segmenting a corresponding one of the voice fragments, and wherein the data of the deterministic component and the data of the stochastic component of each of said voice fragment data each comprise a series of frequency domain data corresponding respectively to the plurality
15 of frames of the frame string corresponding to each of the voice fragments.

 6. A singing voice synthesizing apparatus according to claim 5, wherein said duration time adjusting device generates a frame string of a desired time length by
20 repeating at least one frame of the plurality of frames of the frame string corresponding to each of the voice fragments, or by thinning out a predetermined number of frames of the plurality of frames of the frame string corresponding to each of the voice fragments.

25 7. A singing voice synthesizing apparatus according to claim 6, wherein said duration time adjusting device generates the frame string of a desired time length by repeating a plurality of frames of the frame string corresponding to each of the voice fragments, said
30 duration time adjusting device repeating the plurality of frames in a first direction in which the frame string of a desired time length is generated and in a second direction opposite thereto.

 8. A singing voice synthesizing apparatus according
35 to claim 7, wherein when repeating the plurality of

frames of the frame string corresponding to the data of the stochastic component of each of the voice fragments in the first and second directions, said duration time adjusting device reverses a phase of a phase spectrum of the stochastic component.

9. A singing voice synthesizing apparatus according to claim 1, further comprising a fragment level adjusting device that performs smoothing processing or level adjusting processing on the deterministic component and the stochastic component contained in each of the voice fragment data when the voice fragment data are sequentially concatenated by said synthesizing device.

10. A singing voice synthesizing apparatus according to claim 5, further comprising a deterministic component generating device that changes only pitch of the deterministic component to a desired pitch while preserving the spectral envelope shape of the deterministic component contained in each of the voice fragment data when the voice fragment data are sequentially concatenated by said synthesizing device.

11. A singing voice synthesizing apparatus according to claim 5, wherein said phoneme database stores voice fragment data comprising elongated sounds that are each enunciated by elongating a single phoneme, said phoneme database further storing a flat spectrum as an amplitude spectrum of the stochastic component of each of the voice fragment data comprising each of the elongated sounds, obtained by multiplying the amplitude spectrum thereof by an inverse of a typical spectrum within an interval of the elongated sound.

12. A singing voice synthesizing apparatus according to claim 11, wherein the amplitude spectrum of the stochastic component of each of the voice fragment data comprising each of the elongated sounds is obtained by multiplying an amplitude spectrum of the stochastic

component calculated based on an amplitude spectrum of the deterministic component of the voice fragment data of the elongated sound, by the flat spectrum.

13. A singing voice synthesizing apparatus
- 5 according to claim 12, wherein said phoneme database does not store amplitude spectra of stochastic components of voice fragment data comprising certain elongated sounds, and the flat spectrum stored as an amplitude spectrum of voice fragment data comprising at least one other
- 10 elongated sound is used for synthesis of the certain sounds.

14. A singing voice synthesizing apparatus according to claim 12, wherein the amplitude spectrum of the stochastic component calculated based on the
- 15 amplitude spectrum of the deterministic component has a gain thereof at 0Hz controlled according to a parameter for controlling a degree of huskiness.

15. A singing voice synthesizing method comprising the steps of:
- 20 storing in a phoneme database a plurality of voice fragment data formed of voice fragments each being a single phoneme or a phoneme chain of at least two concatenated phonemes, each of said plurality of voice fragment data comprising data of a deterministic
- 25 component and data of a stochastic component;
- reading out from said phoneme database the voice fragment data corresponding to lyrics inputted by an input device;
- adjusting time duration of the read-out voice
- 30 fragment data so as to match a desired tempo and manner of singing;
- adjusting the deterministic component and the stochastic component of the read-out voice fragment so as to match a desired pitch; and
- 35 synthesizing a singing sound by sequentially

concatenating the voice fragment data that have been adjusted in respect of the time duration and the deterministic component and the stochastic component thereof.

- 5 16. A program for causing a computer to execute a singing voice synthesizing method comprising the steps of:

10 storing in a phoneme database a plurality of voice fragment data formed of voice fragments each being a single phoneme or a phoneme chain of at least two concatenated phonemes, each of said plurality of voice fragment data comprising data of a deterministic component and data of a stochastic component;

15 reading out from said phoneme database the voice fragment data corresponding to lyrics inputted by an input device;

 adjusting time duration of the read-out voice fragment data so as to match a desired tempo and manner of singing;

20 adjusting the deterministic component and the stochastic component of the read-out voice fragment so as to match a desired pitch; and

 synthesizing a singing sound by sequentially concatenating the voice fragment data that have been adjusted in respect of the time duration and the deterministic component and the stochastic component thereof.

- 25 17. A mechanically readable storage medium storing instructions for causing a machine to execute a singing voice synthesizing method comprising the steps of:

30 storing in a phoneme database a plurality of voice fragment data formed of voice fragments each being a single phoneme or a phoneme chain of at least two concatenated phonemes, each of said plurality of voice fragment data comprising data of a deterministic component and data of a stochastic component;

35 reading out from said phoneme database the voice fragment data corresponding to lyrics inputted by an input device;

component and data of a stochastic component;

reading out from said phoneme database the voice fragment data corresponding to lyrics inputted by an input device;

- 5 adjusting time duration of the read-out voice fragment data so as to match a desired tempo and manner of singing;

- adjusting the deterministic component and the stochastic component of the read-out voice fragment so as
10 to match a desired pitch; and

- synthesizing a singing sound by sequentially concatenating the voice fragment data that have been adjusted in respect of the time duration and the deterministic component and the stochastic component
15 thereof.